A Reading-Aware Fusion Fact Reasoning Network for Explainable Fake News Detection

Bofan Wang[1], Shenwu Zhang[2]*
School of Artificial Intelligence, Zhongyuan University of Technology, Zhengzhou, Henan 451191, China[1, 2]
School of Computer Science, Zhongyuan University of Technology, Zhengzhou, Henan 451191, China[1]

***Abstract*—The current growth of information exhibits an exponential trend, with fake news becoming a focal issue for both the public and governments. Existing fact-checking-based fake news detection methods face two challenges: a heavy reliance on fact-checking reports, a lack of explanatory evidence related to the original reports, and a shallow level of feature interaction. To address these challenges, this study proposes a Reading-aware Fusion Fact Reasoning Network for explainable fake news detection. In the aspect of extractive evidence for explainability, a Hierarchical Encoding Layer is constructed to capture sentence-level and document-level feature representations, followed by a Fact Reasoning Layer to obtain report and sentence representations most relevant to the claim, thereby reducing the model's reliance on fact-checking reports. Inspired by reading behaviors, which often involve repeatedly reading the claim and corresponding report during information verification, the Reading-aware Fusion Layer is introduced to learn the deep interdependencies among the claim, evidence, and report feature representations, enhancing semantic integration. Extensive experiments were conducted on the publicly available RAWFC and LIAR fake news datasets. The experimental results demonstrate that RFFR outperforms leading advanced baselines on both datasets.**

***Keywords*—Explainable fake news detection; fact reasoning; feature fusion**

I. INTRODUCTION

The extensive application of the internet, characterized by its convenience and speed, has significantly transformed how individuals acquire and consume information [1]. However, the rapid growth of social media has created an environment conducive to the emergence and swift propagation of fake news [2], leading to severe repercussions and disrupting the balance of authenticity within the news ecosystem [3]. For instance, during the 2016 U.S. presidential election, the most widely circulated fake news stories spread more extensively than the most popular factual news on Facebook. The sheer volume of media content available online makes it exceedingly challenging to manually verify the veracity of news, increasing operational costs for social platforms and hindering early intervention in the dissemination of fake news [4][5].

Existing fake news detection methods can be categorized into two types based on their detection criteria: content pattern-based and fact information-based.

The fake news detection task based on content patterns encompasses two phases:

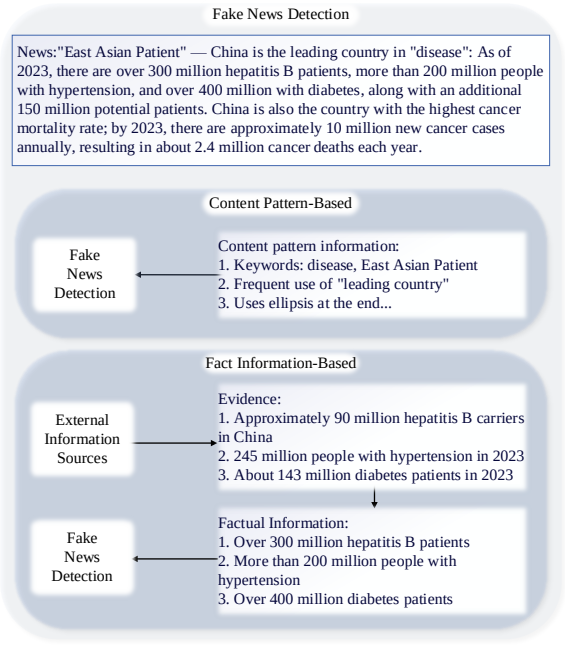


Fig. 1. The news detection process for content pattern-based models and fact information-based models.

1) Unimodal fake news detection. Early studies primarily focused on manually extracting superficial features from textual content, such as punctuation mark frequency [6], and gathering basic platform metadata features [7]. Research then evolved toward the development of neural networks designed to learn semantic features of text [8], emotional characteristics [9], stance-based attributes [10][11], and stylistic features related to the news text content [12]. Additionally, studies explored the use of metadata to capture features based on comments [13] and propagation patterns [14].

2) Multimodal fake news detection. Multimodal detection methods introduce a richer set of distributed feature information into the dataset, including news metadata [15], news images [16], and news videos [17]. Early studies often employed static word vector models for text in conjunction with pre-trained image models to extract semantic features from both text and images [18][19]. They integrated multimodal features to detect fake news through early fusion [20] or late fusion [21] strategies. Recent studies have extensively utilized Transformer pre-trained models to extract advanced semantic features from text and im-

*Corresponding Author

ages [22], capturing semantic similarities across different modalities by establishing entity alignment [23], relationship alignment [24], and semantic alignment mechanisms [25].

While content pattern-based detection methods have achieved satisfactory results in identifying fake news and have reached a relatively mature stage, the knowledge learned by these models is often inductive and based on the training dataset, which may lack global representativeness in its distribution characteristics. News information typically possesses a strong timeliness, with different news items often exhibiting distinct content patterns, making it challenging to generalize to newly emerging news [26]. Consequently, there is growing attention on research related to fake news detection based on factual information. Such methods compare various factual details described in the news, such as event timings, locations, and specific data, with known evidence to discern fake news. Unlike content pattern-based detection, fact-based detection is generally unaffected by writing styles or idiomatic expressions.

As illustrated in Fig. 1, models based on content patterns tend to concentrate on semantics and vocabulary, evaluating whether they conform to patterns indicative of fake news for the same news item. In contrast, models grounded in factual information retrieve evidence related to the news content from external sources, making comprehensive judgments based on the support or lack thereof from the evidence.

Current research on fake news detection utilizing factual information primarily revolves around textual analysis, focusing on two key aspects:

3) Evidence retrieval. The aim of evidence retrieval is to identify high-quality, relevant evidence from fact-checking reports to enhance detection efficacy. Existing methods include search engines [27][28][29], similarity algorithms [30][31], Wikipedia knowledge graphs [32][33], and generative models for evidence production [34].

4) Feature fusion. Current feature fusion methods often employ superficial strategies, such as concatenation, addition, or simple neural networks, to integrate features from different modalities.

However, these methods face several challenges:

In evidence retrieval, there is a heavy reliance on investigative journalism and fact-checking reports that have already been debunked, often neglecting the direct application of original reports. When news has not yet been fact-checked or debunked, many related original reports are usually generated on major media platforms, including media coverage, user comments, and blogs, which typically offer richer factual evidence compared to fact-checking reports [28][35].

In feature fusion, existing methods struggle to capture the internal dependencies between features due to their reliance on superficial fusion strategies.

To address these challenges, a Reading-aware Fusion Fact Reasoning Network (RFFR) has been proposed, focusing on explanatory fact extraction and feature interaction fusion.

In explanatory fact extraction, a Hierarchical Encoding Layer is constructed to capture feature representations of the text, followed by a Fact Reasoning Layer to obtain explanatory evidence. The Hierarchical Encoding Layer consists of a Sentence Encoder and a Document Encoder: the Sentence Encoder captures the hidden features of each sentence in the report, while the Document Encoder provides the hidden representation of the entire report in context. The Fact Reasoning Layer includes a Document Selector to preliminarily screen the top K reports likely containing hidden facts, and a Sentence Selector that refines the most relevant sentences. Throughout this process, we assess whether each sentence in the report constitutes valuable explanatory evidence based on Consistency, Significance, and Redundancy.

In feature interaction fusion, when evaluating the credibility of a claim, individuals typically read the related original reports first and then judge the authenticity of the news based on the extracted evidence [36], often repeating this process. During this time, individuals comprehend the news based on factual evidence while simultaneously understanding the relevant content in the original report. Therefore, there is conditional fusion among claims, evidence, and reports, occurring once or multiple times. Inspired by human reading behavior, a Reading-aware Fusion Layer is constructed to learn the deep dependencies between different feature representations through multiple interactions of claims, evidence, and reports, thereby deepening their semantic integration.

The main contributions of this study are summarized as follows:

- A Hierarchical Encoding Layer is constructed based on original reports to capture sentence and full-text feature representations, followed by a Fact Reasoning Layer that identifies the most relevant report and sentence representations related to the claim, thereby reducing the model's dependence on fact-checking reports.
- Inspired by human reading behavior, a Reading-aware Fusion Layer is proposed to learn dependencies between different features, achieving deep feature fusion.
- Extensive experiments were conducted on two public datasets, RAWFC and LIAR. The results demonstrate that RFFR outperforms other models in fake news detection tasks. Compared to traditional detection models, RFFR consistently achieves superior results across multiple metrics, effectively enhancing the accuracy and performance of fake news identification.

II. Related Work

Based on different criteria for news detection, fake news detection methods can be categorized into content pattern-based and fact information-based approaches.

1) Content pattern-based fake news detection. Significant progress has been made in fake news detection technologies based on content patterns. Early studies mainly relied on manually extracted statistical features, such as the number of punctuation marks [37] and the proportion of negative words [9]. With the development of deep learning technology, methods based on

CNN [38], RNN [13], attention mechanisms [39], and graph architectures [40] have been used to automatically capture semantic, emotional, stylistic, and stance features of text. In addition, social context features, such as metadata of comments [41], user profiles [42], platforms [43], and propagation structures [44], have been used to enhance the detection capabilities of fake news. Subsequently, researchers introduced multimodal data such as images [16] and videos [17], using pre-trained models to obtain semantic features of these data [18][19][20], and combined with different fusion strategies [20][21] and alignment strategies [23][24][25], providing the model with the ability to handle multimodal news.

2) Fact information-based fake news detection. Current research mainly focuses on two aspects:

a) Evidence retrieval: Early studies explored attention mechanisms to highlight significant words [28], news attributes [45], and suspicious users [46], thereby obtaining relevant evidence that provides a certain degree of explanatory support. Later, to improve the readability of word-level methods, techniques such as attention weights [13], semantic matching [47], and entailment [36] were employed to extract evidence sentences. Some studies also obtained explanatory evidence by generating summaries [48] and extracting key points [49]. However, these methods heavily rely on manual fact-checking reports and lack more refined evidence retrieval techniques.

b) Feature fusion: Commonly used feature fusion mechanisms can be roughly divided into two categories: early fusion [50][51][52], also known as feature-level fusion, involves combining information at the feature level through concatenation or addition during the early stages of the model, with the fused features subsequently passed on to downstream learning; late fusion [53][54], or decision-level fusion, relies on the results obtained from each data source individually, integrating them at the final stage of task learning, often using sum, maximum, average, or dot product operations as fusion strategies. However, these methods typically exhibit a superficial degree of feature interaction and fail to uncover hidden associations among different features.

To address these issues, this study presents the Reading-aware Fusion Fact Reasoning Network (RFFR). At the evidence retrieval level, a Hierarchical Encoding Layer is constructed to capture sentence-level and full-text feature representations, followed by a Fact Reasoning Layer (FRL) that identifies the most relevant reports and sentences as explanatory evidence. At the feature fusion level, inspired by human reading behavior, the Reading-aware Fusion Layer (RFL) is proposed and modeled to learn the deep dependencies between different feature representations through multiple interactions of claims, evidence, and reports, which are then applied to the fake news detection task.

III. Problem Statement

Given a fake news dataset $\{D\}$, $D = (c, R)$ is composed of a claim c and a set of related original reports $R = \{r_j\}_{j=1}^{|R|}$, where each $r_i = (s_{i,1}, s_{i,2}, \cdots, s_{i,|r_i|})$ represents a related report composed of a series of sentences. In the fake news detection task, each claim c is associated with a veracity label y, which takes values from $\{True, False, ...\}$, and each original report r_i is associated with a binary label $y_i^r \in Y^r$, indicating whether r_i contains explainable sentences. For each sentence $s_{i,j}$, $y_{i,j}^s \in Y^s$ is a binary label indicating whether $s_{i,j}$ is one of the explainable sentences. The judgment basis for the claim c, denoted as Evidence, is composed of all the explainable sentences. This task can be described as a multi-task learning problem consisting of three sub-tasks: target report selection, explainable sentence extraction, and claim veracity prediction. The goal of this study is to train a model f that satisfies $f(c, D) \rightarrow (\hat{y}, \hat{Y}^d, \hat{Y}^s, \hat{E})$, where $\hat{E}$ represents the explainable judgment basis, composed of a set of predicted sentences (satisfying $\hat{y}_i^d = 1$ and $\hat{y}_{i,j}^s = 1$).

IV. The Proposed Model

The RFFR proposed in this study facilitates explanatory evidence extraction and feature interaction fusion, with its structure illustrated in Fig. 2. It comprises three main modules: the Hierarchical Encoding Layer, the Fact Reasoning Layer, and the Reading-aware Fusion Layer.

A. Hierarchical Encoding Layer

The Hierarchical Encoding Layer is composed of a Sentence Encoder and a Document Encoder. Given a sequence of words $T = (w_1 \cdots w_t \cdots w_{|T|})$ in a claim or report sentence, where $w_t \in R^d$ is a d-dimensional vector initialized by the Sentence Encoder, using the pre-trained language model Bert's final context layer with the special token "[CLS]" as the sentence representation. The sentence representations for each sentence $s_{i,j}$ in the claim c and the original report r_i are $h_c \in R^d$ and $h_{i,j} \in R^d$, respectively.

For document encoding, a Document Encoder based on the self-attention mechanism is constructed. The Document Encoder consists of Multi-Head Attention and a Feed Forward Network through residual connections and Layer Normalization (Add and Norm). The Feed Forward Network is composed of two linear transformations and a ReLU activation function. The output of the self-attention mechanism is passed through a residual connection and then normalized. The core of the Document Encoder is the self-attention mechanism, which is calculated as follows:

$$H = Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where, "Q", "K", *and "*V*"* are the query matrix, key matrix, and value matrix, respectively. Here, $Q = K = V = \hat{h}_i$, $\hat{h}_i = [h_{i,1}; h_{i,2}; \cdots; h_{i,|r_i|}]$, and d_k equals $d/2$. To extensively learn richer contextual information from different perspectives, the multi-head attention mechanism projects the queries, keys, and values m times through different linear projections and executes them in parallel. Finally, the processed results are integrated, projected, and linearly transformed to obtain new representations. The calculation formula for the multi-head attention mechanism is as follows:

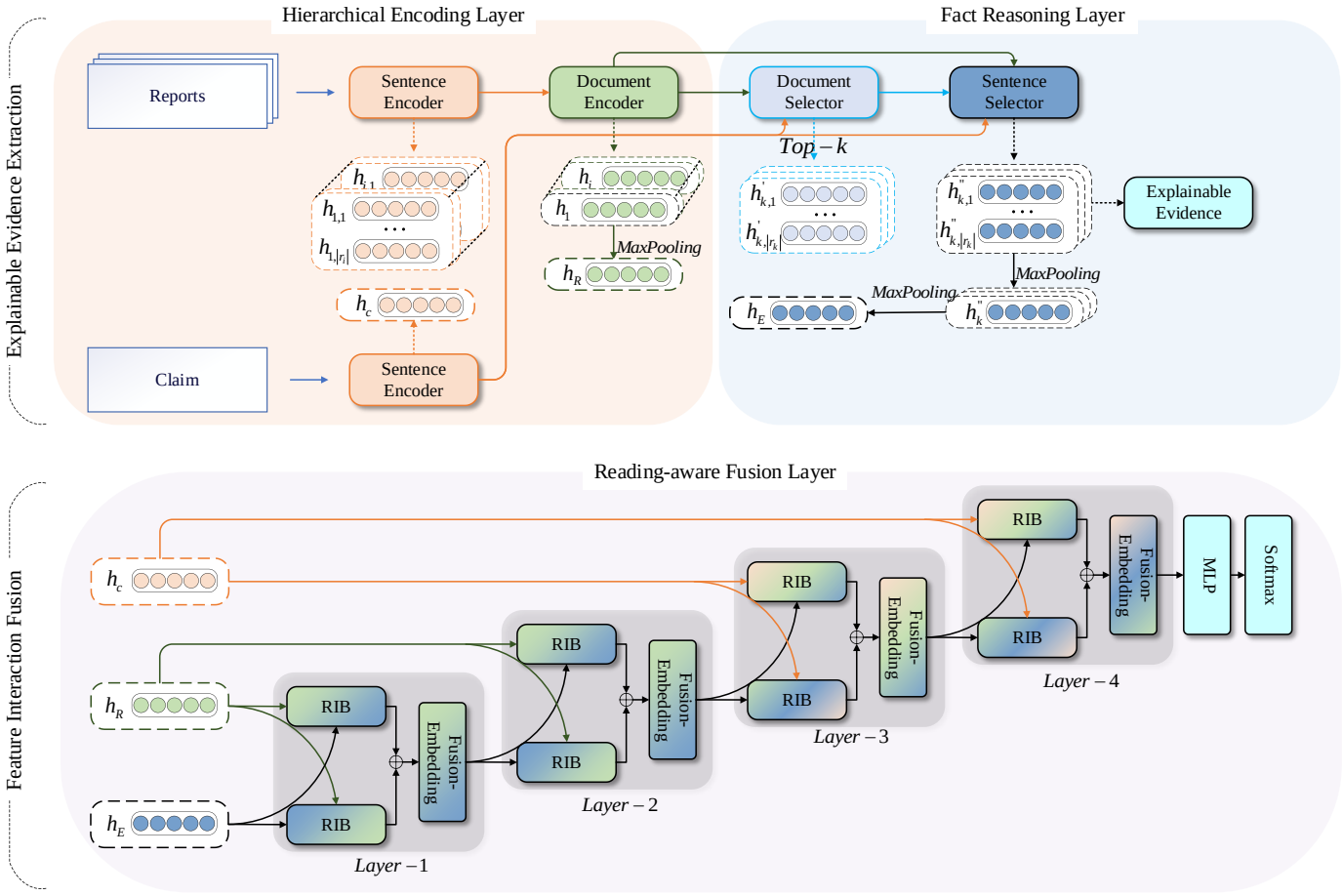


Fig. 2. The proposed RFFR model framework comprises three main layers: the Hierarchical Encoding Layer, which captures sentence-level and full-text representations; the Fact Reasoning Layer, which retrieves relevant reports and sentences as explanatory evidence; and the Reading-aware Fusion Layer, which enhances feature integration and interdependencies for improved detection performance.

$$head = Attention(QW_q, KW_k, VW_v) \quad (2)$$

$$h_i = MultiHead(Q, K, V) = Concat(head_1, \dots, head_m)W_e \quad (3)$$

where, all $W \in \mathbb{R}^{d \times d}$ are trainable parameters, d_e represents the ratio of d to m. $h_i \in \mathbb{R}^d$ is the report representation that integrates all the significant sentence features.

B. Fact Reasoning Layer

To extract explanatory evidence related to the claim from the original reports, this study designs the Fact Reasoning Layer (FRL) in two aspects. First, the Document Selector preliminarily screens the top K reports that may contain hidden facts. Then, a Sentence Selector is designed based on consistency, significance, and redundancy to further refine the most relevant sentences in the original reports.

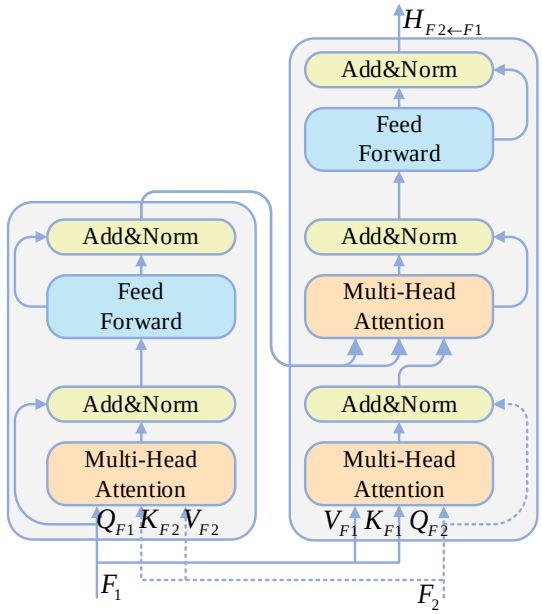
1) Document selector. Since the factual basis corresponding to the claim is hidden in a large number of original reports, the scope of explanatory evidence extraction is automatically narrowed down by ranking the reports and capturing the top-ranked reports. To extract reports that contribute to authenticity prediction from a large number of reports and are worth examining, a coarse-grained Document Selector is developed. The claim is used as the query, and the report is used as the key. An attention weight matrix Att captures the consistency between the claim and the related reports. Att reflects the degree of attention of the claim in the related reports and obtains an importance score for each report r_i.

$$\alpha_c \rightarrow Att = softmax(H_R W_\alpha h_c) \quad (4)$$

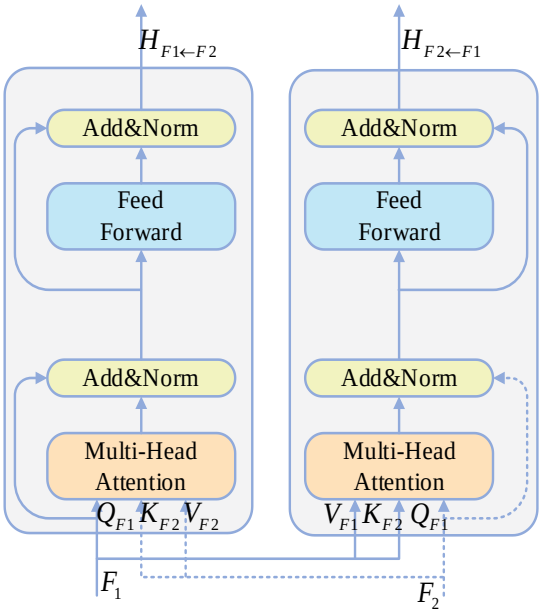
where, $H_R = [h_1; h_2; \cdots; h_{|R|}]$ gathers all the hidden vectors of the reports, and $W_\alpha \in R^{d \times d}$ is a trainable parameter. We use $\alpha_c \rightarrow D$ to rank all reports and select the top K results as reports worth examining. The vector representation of the t-th sentence in the k-th selected report r'_k is represented as $h'_{k,t} \in \{h'_{k,1}, h'_{k,2}, \dots, h'_{k,|r'_k|}\}$, and its document representation is h'_k, which is used for explainable sentence extraction.

2) Sentence selector. Based on the screening of reports, the task of extracting explanatory evidence can be framed as multi-document summarization extraction, where each report is accessed sequentially to identify explainable sentences. Given potential redundancy among reports, multiple original sources are more likely to contain semantically irrelevant and redundant

sentences. This study introduces a fine-grained Sentence Selector, where the selection of explainable sentences is guided by three evaluation indicators.



(a) Reading-aware Interaction Block (RIB)



(b) Traditional co-attention mechanism.

Fig. 3. The architecture diagram of co-attention and our RIB.

Consistency (assessing the relevance of each sentence to the claim theme), Significance (evaluating the importance of each sentence within the report), and Redundancy (determining the novelty and repetition of each sentence relative to previously selected sentences). By integrating these three indicators, the model predicts the probability of each sentence being selected, enabling the extraction of explainable sentences from the original reports.

$$P\left(y_{k,t}^{s}=1\middle|h_c,h'_{k,t},h'_k,h_d\right)= \sigma\left(h'_{k,t}w_c h_c+h'_{k,t}w_r h'_k-h'_{k,t}w_d h_d\right) \quad (5)$$

where, $y_{k,t}^{s}$ is a binary variable indicating whether the t-th sentence in the selected report r'_k should be selected as part of the factual evidence $\hat{E}$, and W_* are trainable parameters. h_d is a redundancy vector initialized to all zeros, updated by the selected sentences in the previously visited reports.

$$h_d=\tanh\left(\sum_t h'_{k-1,t}\cdot P\left(y_{k,t}^{s}=1\right)\right) \quad (6)$$

Considering the number of sentences, RFFR learns to select explainable sentences with selection probabilities higher than the soft threshold $\varepsilon_k=1/|r'_k|$, that is, $P\left(y_{k,t}^{s}=1\right)>\varepsilon_k$. Finally, the integrated representation of the original report and explanatory evidence is obtained through max pooling.

$$h_R=Max\left(\left[h_1;h_2;\cdots;h_{|R|}\right]\right) \quad (7)$$

$$h_E=Max\left(\left[h_1^{"};h_2^{"};\cdots;h_{|K|}^{"}\right]\right) \quad (8)$$

$$h_k^{"}=Max\left(\left[h_{k,1}^{"};h_{k,2}^{"};\cdots;h_{k,|r'_k|}^{"}\right]\right) \quad (9)$$

where, h_R represents the integrated representation of all reports, h_E represents the integrated representation of all explainable sentences, $h''_{k,t}$ represents the sentence representation output from the Sentence Selector, $h_k^{"}$ is the complete representation of explanatory evidence extracted from the k-th report, and K is a hyperparameter controlling the maximum number of selected reports.

C. Reading-Aware Fusion Layer

Intuitively, when reading a report that contains questions, individuals often skim through the article content before focusing on the question section. This process may be repeated multiple times to continuously integrate information from both the question and the report. Based on this reading behavior, this study constructs a Reading-aware Fusion Layer (RFL) to simulate this interaction. The Reading-aware Interaction Block (RIB) serves as the fundamental unit of the RFL, achieving feature fusion by parallelly connecting two RIBs. The RFL is designed to cascade multiple RIB layers for deep feature fusion.

To model human reading behavior, a Reading-aware Interaction Block is developed based on the co-attention mechanism, allowing the model to learn the dependency relationships between different text features, as illustrated in Fig. 3. To achieve comprehensive fusion of text features, four RIB layers are cascaded. The fusion process is gradual, with each RIB layer capable of processing distinct text features, while the output of each layer serves as the input for the subsequent RIB layer. For instance, the input to the first RIB layer is denoted as $<h_E,h_R>$, and its fusion logic is described as follows:

$$H_R=Norm\left(h_R+softmax\left(\frac{h_R h_E}{\sqrt{d}}\right)h_E\right) \quad (10)$$

$$H_E=Norm\left(h_E+softmax\left(\frac{h_E h_R}{\sqrt{d}}\right)h_R\right) \quad (11)$$

$$\widehat{H}_R=Norm\left(H_R+FFN(H_R)\right) \quad (12)$$

$$\hat{h}_{E\leftarrow R}^{(1)}=Norm\left(H_E+softmax\left(\frac{H_E\widehat{H}_R}{\sqrt{d}}\right)\widehat{H}_R\right) \quad (13)$$

$$h_{E\leftarrow R}^{(1)} = Norm\left(\hat{h}_{E\leftarrow R}^{(1)} + FFN\left(\hat{h}_{E\leftarrow R}^{(1)}\right)\right) \quad (14)$$

$$h^{(1)} = h_{E\leftarrow R}^{(1)} \oplus h_{R\leftarrow E}^{(1)} \quad (15)$$

where, $h_{R\leftarrow E}^{(1)}$ is calculated consistently with $h_{E\leftarrow R}^{(1)}$, and $h^{(1)}$ represents the fused semantics of the original report and evidence. $h^{(2)}, h^{(3)}$ are the fused semantics of the corresponding interaction blocks, and $h^{(4)}$ is the final representation of the semantics fusion of claims, evidence, and reports.

D. Learning Task

During the news prediction phase, the final feature representation $h^{(4)}$ is input into a Multi-Layer Perceptron (MLP) layer to predict the authenticity label, as follows:

$$\hat{y} = softmax(\text{MLP}(h^{(4)})) \quad (16)$$

The learning task of the model RFFR is composed of three key sub-tasks: target report selection, explainable sentence extraction, and claim authenticity prediction. Its loss function L_{all} is as follows:

$$L_D = -\sum_i y_i^d \log\left(\hat{y}_i^d\right) \quad (17)$$

$$L_S = -\sum_k \sum_t y_{k,t}^s \log\left(\hat{y}_{k,t}^s\right) \quad (18)$$

$$L_C = -y\log(\hat{y}) \quad (19)$$

$$L_{all} = \theta_D L_D + \theta_S L_S + \theta_C L_C \quad (20)$$

where, L_D, L_S, and L_C represent the cross-entropy losses of the three sub-tasks: target report selection, explainable sentence extraction, and claim authenticity prediction. y_i^d and $\hat{y}_i^d$ represent the true label and predicted label of the report, respectively. $y_{k,t}^s$ and $\hat{y}_{k,t}^s$ represent the true label and predicted probability of the explainable sentence, respectively. y and $\hat{y}$ represent the true label and predicted label of the claim, respectively. β represents the weight parameter, and in the experiment, a multi-task adaptive weighting strategy is used to automatically assign θ_D, θ_S, and θ_C.

1) Multi-task adaptive weighting. To address the parameter optimization problem in multi-task learning, inspired by previous work [55] [56], a Multi-task Adaptive Weighting strategy (MAW) is further proposed to automatically maintain a dynamic balance between tasks on different benchmark datasets. The weighted function $\theta_k(t)$ is defined as follows:

$$\theta_k(t) = \frac{N_k \exp[\gamma_k(t) g(t)]}{\sum_i \exp[\gamma_i(t) g(t)]} \quad (21)$$

$$\gamma_k(t) = \frac{L_k(t-1)}{L_k(t-2)}, g(t) = \frac{\log(t-2)}{T} \quad (22)$$

where, $\theta_k = \theta_k(t)$ represents the weight of the k-th task at the t-th iteration. It indicates the model's focus on task k in the current training step; $N_k = 3$ represents the number of sub-tasks; $\gamma_k(t)$ measures the loss change rate of sub-task k in the last two iterations. If the loss of a task changes significantly, that is, the loss decreases slowly or is unstable, the model will increase the weight of that task; if the loss of a task decreases rapidly, its weight is reduced. $g(t)$ is a global adjustment function used to smooth the weight differences between tasks, making the weight adjustment of different tasks more reasonable and stable; T is a constant value used to control the speed of weight adjustment.

V. Experiments

A. Datasets

This study employs two publicly available datasets to evaluate the proposed method: RAWFC [57] and LIAR [58]. The RAWFC dataset, sourced from Snopes.com, comprises 2012 claims related to fact-checking tasks, each accompanied by supporting evidence and labeled as true, false, or half-true. The LIAR dataset, compiled by PolitiFact.com, includes 12,836 short claims, each categorized as true, mostly true, half-true, barely true, false, or pants-on-fire.

B. Experimental Settings and Evaluation Metrics

To prevent overfitting, the model parameters of Bert were frozen when training on both the RAWFC and LIAR datasets; the hidden dimension in the Sentence and Document Encoders is set to 768; the Document Encoder comprises 12 attention heads and is constructed from 4 attention blocks; when selecting reports, the maximum number of reports K selected for each claim is set to 6 and 9, the soft threshold ε_i is set to $1/|r_i'|$, and the maximum number of explainable sentences extracted is set to 6 and 11 for RAWFC and LIAR, respectively. The batch size is set to 64, the number of training epochs is set to 100, the learning rate is set to 1e-5, and the Adam optimizer is used to minimize the combined cross-entropy loss. To assess the performance of the model, this study uses Accuracy, Macro-averaged Precision (macro-P), Macro-averaged Recall (macro-R), and F1 score (macro-F1) for the detection results evaluation, and ROUGE-N (N=1) and ROUGE-L to assess the quality of the generated explanations.

C. Comparative Baselines and Performance Comparison

To demonstrate the effectiveness of the proposed method, this study compares RFFR with the following baselines:

SVM [59]: Training an SVM-based model for fake news detection using bag-of-words features.

CNN [58]: Integrating available metadata features into representation learning.

RNN [60]: Learning representations from word sequences without external resources.

DeClarE [28]: Combining word embeddings from claims, reports, and sources to assess the credibility of claims.

dEFEND [13]: Using a GRU-based model for veracity prediction and providing explanations.

SentHAN [35]: Representing each sentence based on sentence-level coherence and semantic conflict with the claim.

SBERT-FC [49]: Using SentenceBERT for encoding and detecting fake news based on highly ranked sentences.

GenFE/GenFE-MT [48]: Detecting fake news and providing explanations independently or jointly in a multi-task setting.

TABLE I. COMPARISON RESULTS OF RFFR WITH DIFFERENT BASELINE MODELS ON THESE TWO DATASETS

Model	RAWFC				LIAR			
	Accuracy(%)	P(%)	R(%)	Macro-F1(%)	Accuracy(%)	P(%)	R(%)	Macro-F1(%)
SVM	32.03	32.33	32.51	31.71	14.73	15.78	15.92	15.34
CNN	37.82	38.80	38.50	38.59	21.36	22.58	22.39	21.36
RNN	41.20	41.35	42.09	40.39	21.00	24.36	21.20	20.79
DeClarE	42.18	43.39	43.52	42.18	19.17	22.86	20.55	18.43
dEFEND	42.75	44.93	43.26	44.07	16.64	23.09	18.56	17.51
SentHAN	46.47	45.66	45.54	44.25	19.02	22.64	19.96	18.46
SBERT-FC	46.88	51.06	45.92	45.51	21.75	24.09	22.07	22.19
GenFE	42.66	44.29	44.74	44.43	27.02	28.01	26.16	26.49
GenFE-MT	46.89	45.64	45.27	45.08	15.91	18.55	19.90	15.15
CofCED	50.56	**52.99**	50.99	51.07	28.07	29.48	**29.55**	28.93
RFFR	**52.16**	51.38	**51.76**	**51.13**	**28.39**	**29.68**	29.07	**29.18**

CofCED [57]: A coarse-to-fine cascaded evidence distillation neural network for explanatory fake news detection based on original reports.

TABLE I. presents the detection performance of the proposed RFFR compared to existing strong baselines in terms of precision, recall, and macro F1 (macF1). The following observations can be drawn from the table:

- CNN and RNN outperform SVM on both datasets, indicating that deep learning methods are more effective at capturing semantic and syntactic features from original reports. While dEFEND, DeClarE, and SentHAN achieve better performance on RAWFC by aggregating multiple features from claims, reports, and sources, their results are slightly worse on LIAR. This discrepancy can be attributed to the fine-grained labels in LIAR, which pose additional challenges.
- SBERT-FC and GenFE outperform SentHAN and dEFEND on both datasets, demonstrating the superiority of pre-trained models. Although GenFE-MT performs better than GenFE on RAWFC, it performs significantly worse on LIAR compared to other baselines, highlighting the difficulty of fine-grained fake news detection and explanation generation in a multi-task setting. CofCED achieves better performance than both GenFE and GenFE-MT, suggesting that original reports provide richer explanatory evidence than fact-checking reports.
- Compared to the other models, the proposed RFFR fake news detection model exhibits advantages across various metrics on both the RAWFC and LIAR datasets. Specifically, the accuracy of fake news detection improves by 1.6% on the RAWFC dataset, while the F1 score improves by 0.25% on the LIAR dataset. This enhancement can be attributed to two factors: 1) The Fact Reasoning Layer effectively utilizes different fine-grained selectors to capture more explanatory evidence from original reports, facilitating more accurate fake news detection. 2) The construction of a Reading-aware Fusion Layer for interactive fusion proves more effective than simple feature interaction.

D. Ablation Analysis

1) Effectiveness of each component. To assess the effectiveness of each component in RFFR, eight model variants were created: w/o Report, w/o Evidence, w/o FRL, w/o DS, w/o SS, w/o RFL, w/o RIB, and w/o RFS. These variants indicate the removal of the following components: original report representation, explanatory evidence representation, Fact Reasoning Layer, Document Selector, Sentence Selector, Reading-aware Fusion Layer, Reading-aware Interaction Block, and Reading-aware Fusion Strategy.

The comparison results are shown in TABLE II. yields the following observations:

- RFFR consistently outperforms all ablation experiment variants on both datasets, demonstrating that every component is essential for the model's effectiveness in detecting fake news.
- The w/o Report variant exhibits the poorest performance, followed closely by w/o Evidence. This suggests that Evidence is crucial for the fake news detection task and that the original report provides additional feature information that supports detection, distinct from Evidence. Sole reliance on either the original report or Evidence hinders detection effectiveness.
- The performance of w/o FRL drops significantly, highlighting that noise in the original report adversely affects veracity prediction. The w/o SS variant performs much worse than others, as irrelevant or redundant information in reports can dilute the effectiveness of evidence. The w/o DS variant shows even poorer performance, indicating that noisy reports can impair sentence selection and model training.

TABLE II. ABLATION ANALYSIS ON RAWFC AND LIAR DATASETS

Model	RAWFC				LIAR			
	Accuracy(%)	P(%)	R(%)	Macro-F1(%)	Accuracy(%)	P(%)	R(%)	Macro-F1(%)
w/o Report	32.76	36.76	36.32	35.59	21.16	20.83	20.34	20.24
w/o Evidence	35.59	38.54	38.30	38.07	21.60	22.31	21.52	21.65
w/o FRL	37.88	36.96	40.13	40.45	23.17	23.67	23.04	23.02
w/o DS	40.75	40.12	40.34	39.30	24.31	25.54	24.80	24.06
w/o SS	41.75	40.78	40.85	40.05	22.38	23.71	22.86	23.36
w/o RFL	38.00	38.00	37.83	37.36	23.61	23.92	23.30	23.75
w/o RIB	46.41	45.75	46.0	45.50	25.57	26.41	25.84	25.93
w/o RFS	47.46	46.80	47.10	46.55	26.55	26.99	26.45	26.54
RFFR	**52.16**	**51.38**	**51.76**	**51.13**	**28.39**	**29.68**	**29.07**	**29.18**

- The performance of w/o RFL is significantly reduced, suggesting that simulating human reading behavior facilitates the tight integration of diverse feature information. Both w/o RFS and w/o RIB show substantial performance drops, indicating that these components are vital for deep feature fusion. RIB enables interactive fusion of different features, while RFS enhances the fusion process by mimicking human reading behavior.

2) Comparative analysis of Reading-aware Fusion Layer. The Reading-aware Fusion Layer (RFL) is the mechanism employed by RFFR for deep feature fusion, comprising two core components: the Reading-aware Fusion Strategy (RFS) and the Reading-aware Interaction Block (RIB). The RIB facilitates deep interactive fusion of different features, while the RFS enhances the fusion process by simulating human reading behavior.

Comparative experiments were conducted using traditional Co-Attention [61], Cross-Attention [62], and a variant without RIB (w/o RIB) as alternatives to the RIB, tested under both conditions with and without RFS. The results are presented in the table.

As shown in Fig. 4, the comparison results indicate that the removal of either RIB or RFS significantly impairs the performance of RFFR. In both scenarios (with and without RFS), Cross-Attention, Co-Attention, and RIB consistently outperform the w/o RIB variant, underscoring the necessity of deep interactive fusion among different features. Notably, RIB surpasses both Cross-Attention and Co-Attention, suggesting that RIB achieves more comprehensive feature depth fusion. Furthermore, irrespective of whether alternative methods or RIB are employed, the performance with RFS is consistently superior to that without RFS, highlighting that simulating human reading behavior effectively strengthens the feature fusion process.

3) Comparative analysis of the fact reasoning layer. To examine the impact of the Fact Reasoning Layer (FRL) on explanatory evidence reasoning, three model variants were created: w/o Consistency, w/o Significance, and w/o Redundancy, representing the removal of the respective components. Comparative evaluations were conducted using ROUGE-N (N=1) and ROUGE-L to assess the quality of the explanatory evidence:

LEAD-N [63]: Using the first N sentences as an explanation, where N = 5.

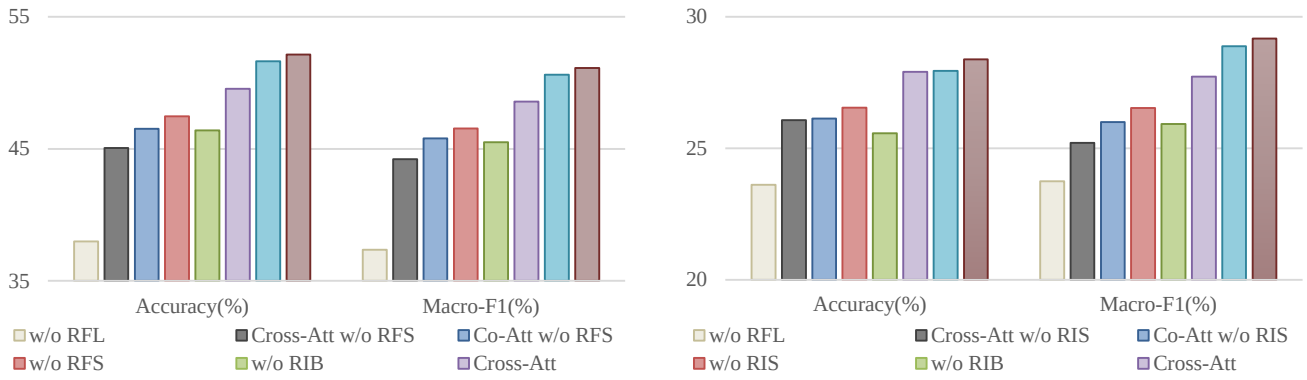
Oracle [48]: Manually selecting the most relevant information from the original report.

dEFEND [13]: Providing explanations based on internal attention weights.

GenFE-MT [48]: Using pre-trained models to generate explanations.

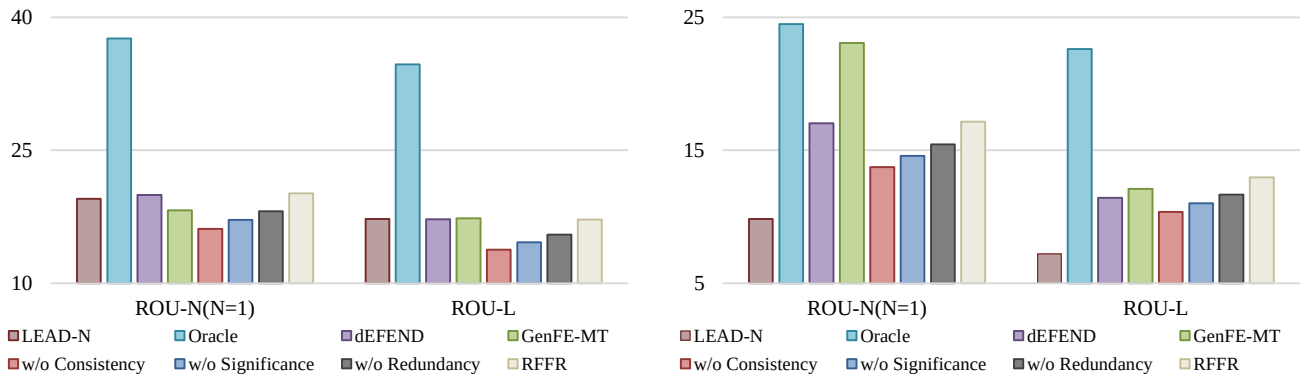
The comparison results, presented in Fig. 5, reveal several key observations: The performance of most models on the LIAR dataset is generally lower than on the RAWFC dataset, indicating that higher granularity in the dataset complicates the generation of explanatory evidence. Additionally, GenFE-MT outperforms dEFEND on both datasets, suggesting that pre-trained models excel in generating explanatory evidence from original reports.

Notably, RFFR outperforms the three ablation variants on both datasets, demonstrating that the evaluations of consistency, significance, and redundancy all enhance the effectiveness of RFFR in generating explanatory evidence. Furthermore, RFFR achieves state-of-the-art performance on RAWFC and displays ROUGE scores closely matching those of GenFE-MT on LIAR, indicating its capability to effectively extract explanatory evidence.



(a) On the RAWFC dataset (b) On the LIAR dataset

Fig. 4. Comparison of performance of different ablation blocks in RFL.



(a) On the RAWFC dataset (b) On the LIAR dataset

Fig. 5. Comparison of performance of different ablation blocks in FRL.

VI. Case Study

To further explore the process of evidence selection in RFFR from original reports, this study selected a case from each of the two datasets and visualized the selection of explanatory sentences, with the results shown in TABLE III. It can be observed that in both the RAWFC and LIAR datasets, sentences with higher scores refuted the claim from various perspectives, while sentences with lower scores contributed less to the prediction of authenticity. This indicates that the assessments of consistency, significance, and redundancy, which are used to select explanatory sentences, are helpful in identifying the main factors for each sentence to serve as explanatory evidence. This method increases the transparency of the system and the credibility of generating explanatory evidence.

VII. Conclusion

Existing fact-checking-based fake news detection methods encounter two main issues: a heavy reliance on fact-checking reports, which often lack explanatory evidence related to the original reports, and a superficial level of feature interaction. To tackle these challenges, this study proposes a Reading-aware Fusion Fact Reasoning Network for explainable fake news detection. At the evidence retrieval level, a Hierarchical Encoding Layer is constructed to capture feature representations of text sentences and the full text, followed by a Fact Reasoning Layer that identifies the most relevant report and sentence representations as explanatory evidence, thus reducing dependence on fact-checking reports. At the feature fusion level, inspired by human reading behavior, a Reading-aware Fusion Layer is introduced to learn dependencies between different feature representations for deep integration. Extensive experiments on the RAWFC and LIAR datasets validate the effectiveness of RFFR. Future work will focus on expanding RFFR for detection by integrating additional data modalities and knowledge bases.

TABLE III. THE PROCESS OF RFFR SELECTING EXPLANATORY EVIDENCE IN THE ORIGINAL REPORT IS VISUALIZED

Dataset: RAWFC **Label:** half **Claim:** U.S. House Speaker Nancy Pelosi publicly criticized the actions of federal agents in U.S. cities during protests in 2020—while simultaneously supporting a budget bill that would fund such law enforcement efforts. **Evidence:** As a comedian who claims he sees the faults in both major U.S. political parties, Jimmy Dore said this to fans amid clashes between Portland, Oregon, protesters and federal agents operating under.....	Consistency	Significance	Redundancy	Overall	Is evidence
1. The new federal bill would change "the standard to evaluate whether law enforcement use of force be justify from whether the force be 'reasonable' to whether the force be 'necessary'" accord to a bill summary send to McClatchy by the office of House Speaker Nancy Pelosi, D-San Francisco.	0.8	0.7	0.9	0.7	1
2. Congressional Democrats announce Monday that they want to raise the legal standard for when law enforcement officer can use deadly force, propose a bill similar to a new California law that aim to reduce lethal encounter.	0.5	0.3	0.4	0.3	0
3. Protests have call attention to the in-custody death of black men and woman, and urge law enforcement reform.	0.2	0.7	0.4	0.3	0
4. The Global Chapter of Black Lives Matter originally support the bill, but later pull it, say amendment to the bill add due to police concern have "significantly weaken" it.	0.8	0.6	0.5	0.8	1
Dataset: LIAR **Label:** pants-fire **Claim:** Hillary (Clinton), one time late at night when she was exhausted, misstated and immediately apologized for it, what happened to her in Bosnia in 1995. **Evidence:** Bill Clinton has implied that the media is biased and covers his wife too harshly. In Boonville, he seemed to lament that unfairness by warning.....	Consistency	Significance	Redundancy	Overall	Is evidence
1. But there be a lot of fulminate because Hillary, one time late at night when she be exhaust, misstate and immediately apologize for it, what happen to her in Bosnia in 1995.	0.3	0.3	0.5	0.2	0
2. And some of them when they 're 60 they 'll forget something when they 're tire at 11 o'clock at night, too.	0.5	0.7	0.7	0.7	1
3. His wife don't make the sniper fire claim one time late at night when she be exhaust.	0.5	0.8	0.3	0.4	0

ACKNOWLEDGMENT

This research was supported by the Key Scientific Research Project of Colleges and Universities in Henan Province, titled "Research on Key Technologies for Text Processing in Network Public Opinion Monitoring" (Project No. 22B520054).

REFERENCES

[1] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media," Big Data, vol. 8, no. 3, pp. 171-188, 2018.

[2] Z. Zhao, P. Resnick, Q. Mei, "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts," in Proceedings of the 24th International Conference on World Wide Web, pp. 1395-1405, 2015.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22-36, 2017.

[4] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, et al., "Detecting rumors from microblogs with recurrent neural networks," In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 3818-3824, 2016.

[5] Y. Zhou, Y. Yang, Q. Ying, Z. Qian, X. Zhang, "Multi-modal Fake News Detection on Social Media via Multi-grained Information Fusion," In Proceedings of the 2023 ACM International Conference on Multimedia Retrieval, pp. 343-352, 2023.

[6] C. Castillo, M. Mendoza, B. Poblete, "Information credibility on twitter," in Proceedings of the 20th International Conference on World Wide Web, pp. 675-684, 2011.

[7] S. Wu, Q. Liu, Y. Liu, L. Wang, T. Tan, "Information Credibility Evaluation on social media," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pp. 4403-4404, 2016.

[8] L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi, et al., "Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge," In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 754-763, 2021.

[9] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, K. Shu, "Mining Dual Emotion for Fake News Detection," In Proceedings of the Web Conference 2021, pp. 3465-3476, 2021.

[10] L. Wu, Y. Rao, H. Jin, A. Nazir, L. Sun, "Different Absorption from the Same Sharing: Sifted Multi-task Learning for Fake News Detection," In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4644-4653, 2019.

[11] J. Xie, S. Liu, R. Liu, Y. Zhang, Y. Zhu, "SERN: Stance Extraction and Reasoning Network for Fake News Detection," In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2520-2524, 2021.

[12] L. Wu, Y. Rao, C. Zhang, Y. Zhao, A. Nazir, "Category-Controlled Encoder-Decoder for Fake News Detection," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 2, pp. 1242-1257, 2023.

[13] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, "dEFEND: Explainable fake news detection," In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 395-405, 2019.

[14] K. Shu, D. Mahudeswaran, S. Wang, H. Liu, "Hierarchical propagation networks for fake news detection: Investigation and exploitation," In Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, no. 1, pp. 626-637, 2020.

[15] Q. Nan, J. Cao, Y. Zhu, Y. Wang, J. Li, "MDFEND: Multi-domain fake news detection," In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3343-3347, 2021.

[16] Y. Zhou, Y. Yang, Q. Ying, Z. Qian, X. Zhang, "Multimodal fake news detection via CLIP-guided learning," In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo, pp. 2825-2830, 2023.

[17] Y. Bu, Q. Sheng, J. Cao, P. Qi, D. Wang, J. Li, "Combating online misinformation videos: Characterization, detection, and future directions," In Proceedings of the 31st ACM International Conference on Multimedia, pp. 8770-8780, 2023.

[18] K. Shu, D. Mahudeswaran, S. Wang, "Hierarchical propagation networks for fake news detection: Investigation and exploitation," In Proceedings of the International AAAI Conference on Web and Social Media, pp. 626-637, 2020.

[19] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, et al., "EANN: Event adversarial neural networks for multi-modal fake news detection," In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 849-857, 2018.

[20] P. Li, X. Sun, H. Yu, Y. Tian, F. Yao, G. Xu, "Entity-Oriented Multi-Modal Alignment and Fusion Network for Fake News Detection," In IEEE Transactions on Multimedia, vol. 24, pp. 3455-3468, 2022.

[21] X. Zhou, J. Wu, R. Zafarani, "Similarity-Aware Multi-modal Fake News Detection," In Proceedings of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 354-367, 2020.

[22] Y. Wu, P. Zhan, Y. Zhang, L. Wang, Z. Xu, "Multimodal Fusion with Co-Attention Networks for Fake News Detection," In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 2560–2569, 2021.

[23] Y. Chen, D. Li, P. Zhang, J. Sui, Q. Lv, et al., "Cross-modal Ambiguity Learning for Multimodal Fake News Detection," In Proceedings of the ACM Web Conference 2022, pp. 2897–2905, 2022.

[24] M. Dhawan, S. Sharma, A. Kadam, R. Sharma, P. Kumaraguru, "Game-on: graph attention network based multimodal fusion for fake news detection," In Social Network Analysis and Mining, vol. 14, pp. 1-13, 2022.

[25] Y. Ganin, V. Lempitsky, "Unsupervised domain adaptation by backpropagation," In Proceedings of the 32nd International Conference on International Conference on Machine Learning, pp. 1180–1189, 2015.

[26] Y. Zhu, Q. Sheng, J. Cao, S. Li, D. Wang, F. Zhuang, "Generalizing to the Future: Mitigating Entity Bias in Fake News Detection," In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2120–2125, 2022.

[27] K. Popat, S. Mukherjee, J. Strötgen, G. Weikum, "Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media," In Proceedings of the 26th International Conference on World Wide Web Companion, pp. 1003–1012, 2017.

[28] K. Popat, S. Mukherjee, A. Yates, G. Weikum, "DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning," In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 22–32, 2018.

[29] Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, J. G. Simonsen, et al., "MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking of Claims," In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4685-4697, 2019.

[30] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, "FEVER: a Large-scale Dataset for Fact Extraction and VERification," In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 809-819, 2018.

[31] B. M. Yao, A. Shah, L. Sun, J. Cho, L. Huang, "End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models," In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2733-2743, 2023.

[32] J. Kim, S. Park, Y. Kwon, Y. Jo, J. Thorne, E. Choi, "FactKG: Fact Verification via Reasoning on Knowledge Graphs," In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 16190-16206, 2023.

[33] Zou, Z. Zhang, H. Zhao, "Decker: Double Check with Heterogeneous Knowledge for Commonsense Fact Verification," In Findings of the Association for Computational Linguistics: ACL 2023, pp. 11891-11904, 2023.

[34] P. Nakov, D. Corney, M. Hasanain, F. Alam, T. Elsayed, A. Barrón-Cedeño, et al., "Automated Fact-Checking for Assisting Human Fact-Checkers," In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, pp. 4551-4558, 2021.

[35] J. Ma, W. Gao, S. Joty, K.-F. Wong, "Sentence-Level Evidence Embedding for Claim Verification with Hierarchical Attention Networks," In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 2561-2571, 2019.

[36] Y. Wang, S. Qian, J. Hu, Q. Fang, C. Xu, et al., "Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks," In Proceedings of the 2020 International Conference on Multimedia Retrieval, pp. 540-547, 2020.

[37] S. B. Parikh, P. K. Atrey, "Media-Rich Fake News Detection: A Survey," In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval, pp. 436-441, 2018.

[38] P. K. Verma, P. Agrawal, I. Amorim, R. Prodan, "WELFake: Word Embedding Over Linguistic Features for Fake News Detection," IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 881-893, 2021.

[39] L. Wu, Y. Rao, Y. Zhao, H. Liang, A. Nazir, "DTCA: Decision Tree-based Co-Attention Networks for Explainable Claim Verification," In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 1024-1035, 2020.

[40] L. Hu, T. Yang, L. Zhang, W. Zhong, D. Tang, C. Shi, et al., "Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge," In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 754-763, 2021.

[41] H. Choi, Y. Ko, "Using Topic Modeling and Adversarial Neural Networks for Fake News Video Detection," In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 2950-2954, 2021.

[42] Y. Dou, K. Shu, C. Xia, P. S. Yu, L. Sun, "User Preference-aware Fake News Detection," In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2051-2055, 2021.

[43] P. Qi, J. Cao, T. Yang, J. Guo, J. Li, "Exploiting Multi-domain Visual Information for Fake News Detection," In 2019 IEEE International Conference on Data Mining, pp. 518-527, 2019.

[44] K. Shu, D. Mahudeswaran, S. Wang, H. Liu, "Hierarchical Propagation Networks for Fake News Detection: Investigation and Exploitation," In Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 626-637, 2020.

[45] Y. Liu, M. Lapata, "Text Summarization with Pretrained Encoders," In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 3730-3740, 2019.

[46] Y.-J. Lu, C.-T. Li, "GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media," In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 505-514, 2020.

[47] Y. Nie, H. Chen, M. Bansal, "Combining fact extraction and verification with neural semantic matching networks," In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, pp. 842-849, 2019.

[48] P. Atanasova, J. G. Simonsen, C. Lioma, I. Augenstein, "Generating Fact Checking Explanations," In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7352-7364, 2020.

[49] N. Kotonya, F. Toni, "Explainable Automated Fact-Checking for Public Health Claims," In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 7740-7754, 2020.

[50] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, S. Satoh, "SpotFake: A Multi-modal Framework for Fake News Detection," In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data, pp. 39-47, 2019.

[51] S. Y. Boulahia, A. Amamra, M. R. Madi, S. Daikh, "Early Intermediate and Late Fusion Strategies for Robust Deep Learning-based Multimodal Action Recognition," Machine Vision and Applications, vol. 32, no. 6, pp. 1-18, 2021.

[52] J. Xue, Y. Wang, Y. Tian, Y. Li, L. Shi, L. Wei, "Detecting fake news by exploring the consistency of multimodal data," Information Processing & Management, vol. 58, no. 5, pp. 102610, 2021.

[53] P. Meel, D. K. Vishwakarma, "Multi-modal fusion using Fine-tuned Self-attention and transfer learning for veracity analysis of web information," Expert Syst. Appl., vol. 229, pp. 1-16, 2023.

[54] S. Singhal, M. Dhawan, R. R. Shah, P. Kumaraguru, "Inter-modality Discordance for Multimodal Fake News Detection," In Proceedings of the 3rd ACM International Conference on Multimedia in Asia, pp. 33-39, 2022.

[55] Z. Chen, V. Badrinarayanan, C.-Y. Lee, A. Rabinovich, "GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks," In Proceedings of the 35th International Conference on Machine Learning, vol. 80, pp. 794-803, 2018.

[56] S. Liu, E. Johns, A. J. Davison, "End-To-End Multi-Task Learning With Attention," In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1871-1880, 2019.

[57] Z. Yang, J. Ma, H. Chen, H. Lin, Z. Luo, Y. Chang, "A Coarse-to-fine Cascaded Evidence-Distillation Neural Network for Explainable Fake News Detection," In Proceedings of the 29th International Conference on Computational Linguistics, pp. 2608-2621, 2022.

[58] W. Y. Wang, "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 422-426, 2017.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825-2830, 2011.

[60] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, Y. Choi, "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931-2937, 2017.

[61] Lu, J., Batra, D., Parikh, D., Lee, S. "ViLBERT: Pretraining TaskAgnostic Visiolinguistic Representations for Vision-and-Language Tasks," In Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 2-12, 2019.

[62] Chen, C.-F. R., Fan, Q., Panda, R. "CrossViT: Cross-Attention MultiScale Vision Transformer for Image Classification," In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, pp. 347-356, 2021.

[63] R. Nallapati, F. Zhai, B. Zhou, "SummaRuNNer: a recurrent neural network based sequence model for extractive summarization of documents," In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 3075-3081, 2017.